

電子データをどのような形で後世に残すか ー総合情報基盤センター・デジタル・アーカイブス構築奮戦記ー

総合情報基盤センター 教授 高井 正三

“電子データ”をどのような保存形式で、どのような記録媒体を使って後世に残すか。これは電子社会の長年の課題であり、未だ正解が出ていない様である。筆者は現在、当時「富山大学計算センター」が設置されてから、50年間に蓄積された膨大な文献・資料の中で、歴史に残して、後世に伝えるべきものを選択し、PDF形式に電子化して、Web「総合情報基盤センター・デジタル・アーカイブス ITC Digital Archives」上に Upload している。Web上に掲載し、後輩諸君に、貴重な文献・資料を追加し、公開作業を継続していけば、今年の「新生富山大学10周年」を100年以上は継続できるであろう。同時に「富山大学デジタル・アーカイブス」も構築し、継続して欲しい。本稿では、電子データの永年保存に向けた提案、ITC Digital Archives 構築奮戦記を紹介する。

1. 電子データ永年保存の課題

1.1 記録媒体の寿命と保存

電子データ”，あるいは記録媒体上の“デジタル・データ”を，どのような形＝保存形式で後世に残すか。長年と言っても42年間であるが，いつも考えてきた。図1.1は毎年富山県児童クラブ連合会の指導者研修会で，講演の最初に問いかけるクイズである。

クイズ3: 次の数字は何を意味しているでしょう?

- 紙(A)250~700年, 紙(B) 50~150年
- マイクロ・フィルム 500年
- 映画フィルム(100年/TACベース, 500年/PETベース)
 - TAC(tri-acetyl cellulose), PET(polyethylene terephthalate)
- LPレコード(Long Playing Record) 100年
- 磁気テープ, ビデオ・テープ 30年
- フロッピー・ディスク 20年
- 音楽CD(Compact Disc), CD-ROM 30年
- CD-R 10~30年
- DVD-ROM 30年, DVD-R 20年
- USBメモリー 4/5年~10年(データ保持期間)
- ブルーレイ・ディスク(Blu-ray Disc) 20~30, ~Max.50年
- M-DISC(Millennial Disc: 千年円盤という意味) 100年
- HDD(Hard Disk Drive) 5年(耐用年数)

2014/11/16 Information Technology Center, University of Toyama 3

図 1.1 記録媒体の寿命（和紙が一番長い）

結論から言えば，和紙に墨で記録し，正倉院のような環境下で保存すると，和紙が記録媒体として一番寿命が長いようだ。次に推奨するのは PET(polyethylene terephthalate)ベースの映画フィルムまたは Microfilm, Microfiche で，寿命は500年とされている。購読していた COLLECTED ALGORITHMS FROM ACM は2冊目まで紙媒体であったが，3冊目から Microfiche の形式に変更され，1枚の Microfiche に300ページ分のアルゴリズムが記載されるようになった（写真1.2）。

以上はアナログ形式の記録媒体であるが，デジタル時代の記録媒体で一番寿命が長いのは Blu-ray Disc か M-DISC か。それとも MRAM (Magnetoresistive Random Access Memory) の量産化が進めば，MRAM は電荷ではなく磁気モーメントによって情報を記憶するため，リークや摩耗の心配がなく，無限の書き換えサイクルが可能となり，大いに期待できる。

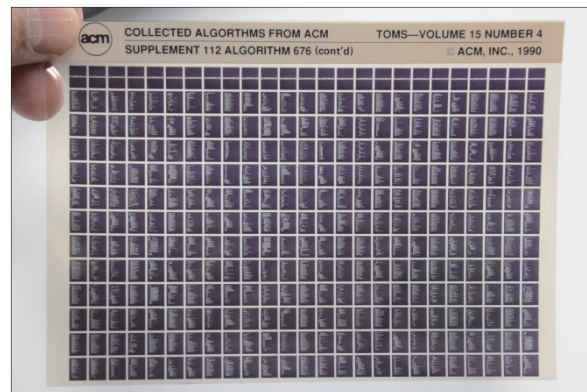


写真 1.2 マイクロフィッシュ Microfiche の画像

1.2 国立国会図書館の場合

この記録媒体の寿命と保存に感心があるのは，資料のデジタル化と電子情報の保存を推進する我が国の国立国会図書館（National Diet Library : NDL）も同じで，NDL のホームページ「よくある質問」のコーナーに「資料の保存・デジタル化」サイトで，ユーザーの質問に答えている（図1.3）。

質問の最後に「Q:電子情報の保存に関して，先進的な取り組みを教えてください。」の回答は，「A:海外，特に欧米では，各国の図書館や公文書館，学術機関等が連携している先進的

なプロジェクトが多くあります。たとえば、米国議会図書館 (LC) が主導する全米デジタル情報基盤整備・保存プログラム (National Digital Information Infrastructure and Preservation Program; NDIIPP) や英国のデジタル保存連合 (Digital Preservation Coalition; DPC) 等が、電子情報の長期的な保存に関する調査研究や課題解決等に活発に取り組んでいます。」である。この米国、英国に刺激されてか、国立国会図書館の資料デジタル化に係る基本方針が次のサイトに掲載されている。

【URL=<http://www.ndl.go.jp/jp/aboutus/digitization/policy.html>】



図 1.3 NDL よくある質問：電子情報の保存 Page

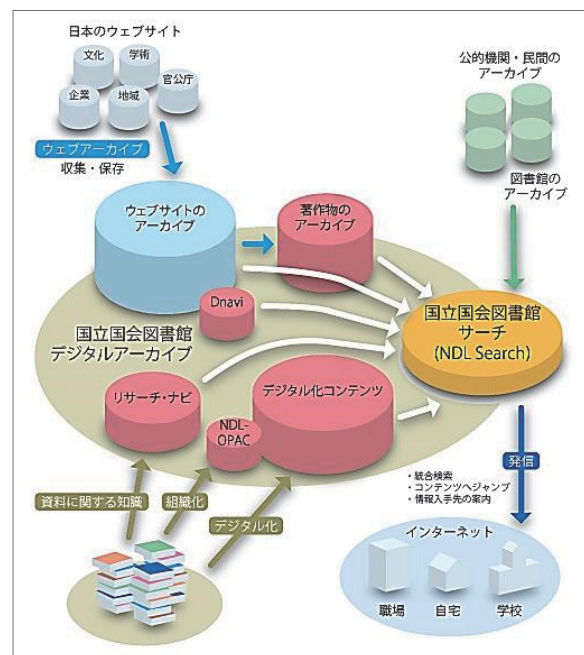


図 1.4 NDL デジタル・アーカイブの全体イメージ

さらに、NDL デジタル・アーカイブ・システムについて、全体イメージ図を掲げている (図 1.4)。

2. 記録媒体への保存形式 (Format)

筆者は、これらの記録媒体へのデジタル保存形式について、以下の様に提案したい。

(1) 文書・資料、画像の保存形式

文書類や画像は、最も普及している画像圧縮形式 JPEG か、Adobe Systems 社が開発した PDF (Portable Document Format) がベターと考えている。これは現在使用しているイメージ・スキャナーの出力形式に採用されていることによる。JPEG⇔PDF の変換も容易にできるのも良い。

(2) 音響・映像の保存形式

音響に関しては、SONY と Phillips が共同で開発した、音楽 CD 規格 CD-DA (Compact Disc Digital Audio) から、これをよりコンパクトにした音響ファイル圧縮形式の MP3 (MPEG Audio Layer-3) か、電子楽器の演奏データを機器間でデジタル転送するための世界共通規格 MIDI (Musical Instrument Digital Interface) がベターのようである。

映像の保存形式は Analog の映画フィルムには及ばないが、MPEG-4 (Moving Picture Experts Group -4) が、メーカーに依存しないので推奨したい Format である。

(3) 数値データの保存形式

数値データは、過去から現在までを網羅する意味で IBM および IEEE745 形式を併用しても良いが、PC の普及に伴って IEEE 形式が標準になってきた。

(4) テキスト・データの保存形式

テキスト・データは、Unicode の UTF-8 を使用するのが一般的である。古文書から最新資料 (青空文庫のようなテキスト) まで、すべての文字コードを UTF-8 にして、標準化を図るべきである。MS の Shift-JIS コードが未だに幅を効かしているのが不思議である。最近では絵文字まで登録されているが、100 年後に使われているだろうか心配ではある。

3. ITC Digital Archives 奮戦記

3.1 スキャナーでの紙詰まり対策

古い文献・資料をスキャナーで読み込む場合が殆どで、度々紙詰まりが発生し、貴重な文献・資料が「しわくちや」の状態に陥り、それを伸ばして＝前の状態にして、再度、正常に読み込むのが大仕事である。正常に読み込まない原因は、

(1) 冊子体の場合、無線綴じくみ製本のノド側＝背を裁断用カッターで 1～2 mm 裁断するのだが、総てのページがきれいに裁断されているかを、「ばらばらっと」裁いてみるだけで、2～3 枚が糊でくっついていいる場合が多々あり、これが紙詰まり原因になる。

(2) 用紙が古い場合、「藁半紙(わらばんし)」と言った時代の、スキャナーも恐れそうな紙質で、一度読み込んだら、破れそうなものも多い。これが正常にスキャナーに読み込まれない場合があり、紙詰まりを起こす。こういう場合は、フラット・ベッド型のスキャナーでスキャンした方が良い。

(3) ホチキス留めの資料でも、裁いていない場合は 2～3 枚同時に読み込み、機械的に資料が引き込まれて止まり、「しわくちや」の状態になる。これが心配になるなら、時間を惜しまず、1 枚ずつスキャナーに入れてやることである。

3.2 スキャン作業と設備

スキャン作業は、スキャナーなど設備を一式揃え、使いこなす中で、様々な問題を工夫しながら解決していくより方法が無いようだ。



写真 3.1 ScanSnap iX500 でスキャンしている所

写真 3.1 は、筆者の研究室の作業場である。昨年調達した 3 代目のスキャナー FUJITSU ScanSnap iX500 と、左下がスキャナー制御と

データ保存用の Windows Desktop Computer である。机にはカッター・ナイフ用のカッティング・マットを敷いている。何時でもカッター・ナイフを使えるようにしている。写真 3.2 は、パンフレットや冊子など裁断せずに、見開きのままスキャンし、歪みを補正してくれるスキャナー FUJITSU ScanSnap SV600 である。



写真 3.2 ScanSnap SV600 でスキャンしている所
写真 3.3 は、平綴じや無線綴じなどの製本の背(表紙)部分を、1～2 mm 裁断するのに使用しているカッター PLUS PK-513L である。よく平綴じに使われている針金もろとも裁断するため、数カ所刃こぼれが目立ってきたようだ。



写真 3.3 背を裁断するカッター PLUS PK-513L

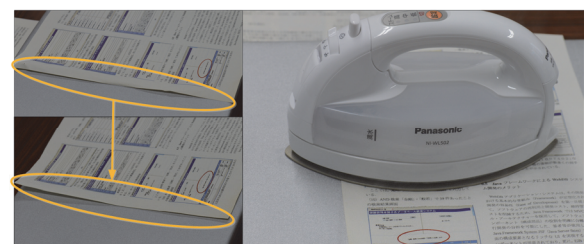


写真 3.4 しわを伸ばすためのスチーム・アイロン

写真 3.4 は、「しわくちや」になった資料を伸ばすためのスチーム・アイロン(右)と、しわになった文書を伸ばした時の前後写真(左)である。スキャナー読み取り時に、「しわくちや」になった用紙を元通りに復元するためには、十分な水分(スチーム)を与えてから、表と裏の両面で伸ばさないと、元の状態に復元するのは難しいようだ。

3.3 スキャナーでの歪みの補正

ScanSnap SV600 では、本を裁断しなくてもスキャンして、読み込み時の歪みを補正できるようになった。先ず ScanSnap Manager の設定画面で図 3.5 のように SV600 を選択し、原稿とサイズの選択を行い、写真 3.2 のように、総てのページをスキャンします。読み取りを終えたら、[読み取り終了] ボタンをクリックします(図 3.6)。



図 3.5 Scanner／原稿／サイズの選択



図 3.6 「読み取り終了」ボタンをクリック

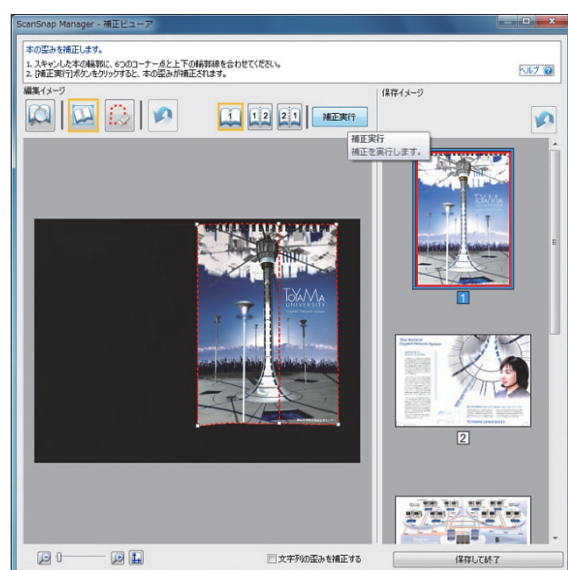


図 3.7 補正ビュー・ダイログ・ボックス
総てのページを読み込み終了後に、図 3.7 の

様な ScanSnap Manager の補正ビューアが、「本の歪みを補正します。」と表示してくるので [補正実行] ボタンを押下すると、図 3.8 のダイアログ・ボックスが表示され、「対象ページを選択して下さい。」と聴いてくるので、「すべてのページ」または「選択しているページ」を選ぶと、図 3.9 の左側を右側の様に補正してくれる。

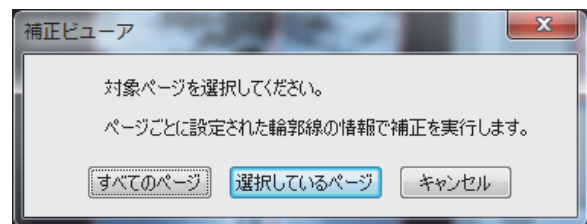


図 3.8 補正ビューア選択ダイログ・ボックス

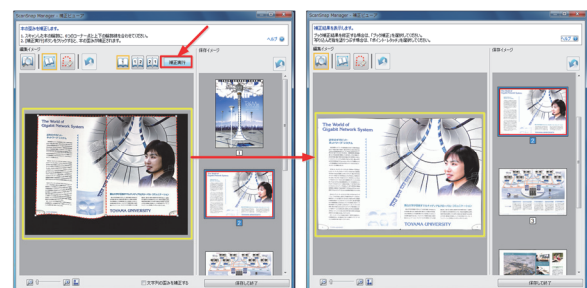


図 3.9 ScanSnap Manager の補正ビューア

3.4 情報の分類・選択と歴史的遺産の保存

最終的に、スキャンした文献・資料は“ITC Digital Archives”に保存するが、公開する情報は、センター年代別、項目別に分類し、Web 上で公開する資料は、著作権処理と情報の公開基準に従って選択する必要がある。また、貴重資料の中には歴史的遺産も多々含まれているので、これをどのように保存していくかも、充分吟味して、後世に伝えていかなければならない。

4. むすび

“ITC Digital Archives”という電子保管庫は、“University of Toyama Digital Archives”として拡充・発展させ、歴史的遺産を後世に伝えるために、引き継いでいってもらいたい。何度も主張するが、「その資料がなくてはならないなら、それは伝えるために歴史に残せ！」[1]と。

参考文献

[1] “総合情報基盤センター・デジタル・アーカイブスの開設について”，高井正三，総合情報基盤センター広報，Vol.11，45-50，2014。